Modelling Temperature Dynamics

Master Thesis Submitted to

Prof. Dr. Brenda López Cabrera

Prof. Dr. Wolfgang K. Härdle

Ladislaus von Bortkiewicz Chair of Statistics

C.A.S.E.- Centre for Applied Statistics and Economics

Humboldt-Universität zu Berlin

by

Zografia Anastasiadou

(522088)

in partial fulfillment of the requirements

for the degree of

Master of Science in Statistics

Berlin, 11 January 2012

Acknowledgment

I would like to express my deep gratitude to my supervisor, Prof. Dr. Brenda López Cabrera, for her help, advice and encouragement throughout this work. I would also like to deeply thank my advisor, Prof. Dr. Wolfgang K. Härdle, for his constant support and insightful comments in all the time of writing my thesis.

In addition, I would like to extend my thanks to Mengmeng Guo, whose critical comments and help in understanding temperature data was vital to me. Also I would like to thank the whole staff of the Ladislaus von Bortkiewicz Chair of Statistics for constructive criticism.

Last but not the least, I would like to thank my parents, Fotini and Stergios Anastasiadis, for supporting me emotional throughout my life. Their understanding and encouragement were essential for me to accomplish this work.

Zografia Anastasiadou

Abstract

This work presents a time series model for daily average temperatures. The data is modeled by flexible low-order autoregressive terms, seasonality components and a deterministic volatility capturing the heteroscedasticity of the residuals. The study attempts to find evidence in shifts in the variance part over time which is attributed to the global warming effect. The model is applied to the industrial Bleu Banana European area with the data covering the period from 1973 to 2008. After deseasonalizing and detrending the data, four standard approaches for modelling daily temperature dynamics are estimated and evaluated. We found out that the multiplicative model of Fourier and GARCH terms in volatility outperforms the others. Furthermore, expectile curves and quantile curves, applied on temperature residuals, are presented to detect fluctuations in temperature variance and evidence for global warming.

Keywords: Global Warming, Temperature Data, Seasonality, Autoregressive Model, Expectile Curves, Quantile Curves

Zusammenfassung

Gegenstand dieser Arbeit ist ein Zeitreihenmodell für tägliche Durchschnittstemperaturen. Die Daten werden mit Hilfe der flexiblen Klasse von Autoregressiven Prozessen geringer Ordnung und einer Saisonkomponente modelliert, wobei die Heteroskedastizität in den Residuen über einen deterministischen Volatilitätsparameter modelliert wird. Es wird nach Anzeichen für eine Verschiebung im Varianzteil gesucht, die auf die globale Erwärmung zurückgeführt werden kann. Die Daten für das Modell stammen aus dem industriellen Gebiet der sogenannten Blauen Banane und wurden im Zeitraum von 1973 bis 2008 erhoben. Nach dem Entfernen von Trends und Saisonalität in den Daten werden diese benutzt um vier dynamische Standardmodelle für Tagestemperaturen zu schätzen und auszuwerten. Es wird gezeigt, dass das multiplikative Modell mit Fourier- und GARCH-Termen in der Volatilitätsstuktur eine bessere Leistung als die anderen Standardmodelle erzielt. Des Weiteren untersuchen wir die Expektil- und Quantilkurven der Temperaturresiduen um Fluktuationen in der Varianz der Temperatur nachzuweisen, welche ein Anzeichen für globale Erwärmung ist.

Schlagwörter: Globale Erwärmung, Temperatur Datensätze, Saisonalität, Autoregressives Modell, Expektilkurven, Quantilkurven

Contents

List of Figures

List of Tables

1 Introduction

Modelling and forecasting of weather data is a challenging task in agricultural and financial economics. The real economy is highly influenced by the climate and weather conditions. In particular the energy consumption and demand, the agricultural industry and the tourism sector are significantly affected by the average temperature. Furthermore, pricing of weather derivatives, is a task of increasing importance. Therefore a vast literature has been devoted to study the weather dynamics and global warming. In this thesis we focus our attention on modelling the evolution of the daily average surface air temperatures. Many researchers have focused on finding shifts in the mean of daily average temperatures, but only a few have done this for the variance. In this work we extend our analysis in detecting shifts in variance of daily average temperatures. We found out that the data volatility changed over recent 36 years for the industrial Blue Banana European area. These changes provide evidences and are attributed to the global warming effect. Moreover, those changes can help us understand the occurrence of extreme weather events.

The temperature dynamics were investigated by many prominent researchers. Campbell and Diebold (2005) proposed an autoregressive process of higher order to capture the dynamics of the deseasonalized temperature and detected a seasonal pattern in the temperature variation. The model suggested by Benth et al. (2007) and Benth et al. (2011) incorporates a deterministic seasonal component and a higher order continous autoregressive process with seasonal variation. In those models the seasonal component consists of higher order of polynomials and trygonometric functions. To relax this restriction we followed Song et al. (2010) approach. The selection of the seasonal components, the upward trend, seasonal effect and solar activity impact, is done by Lasso technique and is compared with the local linear regression, see Härdle et al. (2011a). Therefore overfitting issues are avoided and the impact of the seasonal pattern in the data is controled. The volatility of the detrended and deseasonalised time series is modeled by four different approaches: the Local Linear Regression (LLR), Härdle et al. (2011a), a model of Fourier series proposed by Benth et al. (2007), an additive model of Fourier and GARCH terms, Campbell and Diebold (2005) and a multiplicative model of Fourier and GARCH terms, Benth and Saltyte Benth (2011). The performance of the models is evaluated using residual analysis. Secondly, we analyse the changes in the temperature variance by expectile (Guo and Härdle (2012)) and quantile regression (Härdle et al. (2011b)), estimated for different time periods.

This master thesis is structured as follows. In section 2 we introduce the Temperature Dynamic Model with seasonality and inter temporal autocorrelation. Four different models are shown for the seasonal volatility of the deseasonalised and detrended data. Statistical tests are presented for the standardized data by the estimated seasonal volatilities. In Section 3 the models, proposed in Section 2, are applied to the temperature data of four European cities, London, Rome, Paris and Amsterdam. In section 4 we evaluate the performance of the multiplicative model of Fourier and GARCH terms. Section 5 includes the results of the expectile regression. The quantile regression results are presented in Section 6. Section 7 concludes. The computations

for sections 3, 4, 5 and 6 were done with R version 2.12.2 and Matlab version 7.10.0.

2 The Temperature Dynamic Model

In order to estimate the evolution of temperature in time, the following discrete model for temperature dynamics as in Benth et al. (2011), is constructed

$$T_t = X_t + \Lambda_t \tag{2.1}$$

- T_t is the average temperature in day t, $t = 1,...,M$. T_t is computed as $T_t = \frac{T_{t,max}+T_{t,min}}{2}$.
- Λ_t is the seasonal function which is approximated with two estimation techniques, the Local Linear Regression (LLR) estimator, as proposed by Härdle et al. (2011a):

$$\arg\min_{e,f} \sum_{t=1}^{365} \{\bar{T}_t - e_s - f_s(t-s)\}^2 K\left(\frac{t-s}{h}\right) \tag{2.2}$$

where $\bar{T}_t$ is the mean over years of daily averages temperatures, h is the bandwidth, $K(\cdot)$ a Kernel and the least squares seasonal function:

$$\Lambda_t = \alpha + \beta t + \gamma(3t^2 - 1) + \sum_{i=1}^{K} \delta_i cos\{2\pi i(t-\zeta_i)/365\} + \sum_{i=1}^{K} \kappa_i cos\{4\pi i(t-\xi_i)/365 \cdot 11\} \tag{2.3}$$

We use the Lasso algorithm, proposed by Tibshirani (1996) for the selection of the coefficients of equation (2.3). The Lasso algorithm has the form

$$\arg\min_{\beta} \left(\|T_t - Y\beta\|^2\right) + \lambda\|\beta\|_1 \tag{2.4}$$

where T_t is the vector of daily averages temperatures, Y is the vector of time basis (the explanatory variables), $\|\cdot\|_1$ denotes the L_1-norm and λ is a penalty term which shrinks the coefficients β, in our case the coefficients of equation (2.3) to zero.

Next, we remove the seasonality from the daily average temperatures and before we model the remaining residuals X_t with an autoregressive process we check whether X_t is stationary. For that reason two tests are applied, the Augmented Dickey-Fuller (ADF) test for testing the unit root hypothesis and the KPSS test for testing the stationarity hypothesis.
The ADF test is of the form

$$(1-L)X = c_1 + \mu t + \tau LX + \alpha_1(1-L)LX + \ldots + \alpha_p(1-L)L^pX + \varepsilon_t, \tag{2.5}$$

where p is the lag order of the autoregressive process. The null hypothesis of the test is $H_0 : \tau = 0$ (unit root) versus the alternative $H_1 : \tau < 0$ (stationarity) by means of the test statistic of the OLS estimator of τ.

The regression model of the KPSS test has the form

$$X_t = c + \mu t + k \sum_{i=1}^{t} \xi_i + \varepsilon_t, \tag{2.6}$$

where ε_t is stationary and ξ_t is i.i.d with an expected value 0 and variance 1. The null hypothesis is $H_0 : k = 0$ while the alternative hypothesis is $H_1 : k \neq 0$.

If H_0 of ADF test is rejected and the H_0 of KPSS test cannot be rejected then X_t is a stationary process and can be modeled with an autoregressive process $AR(p)$, defined as:

$$X_{t+p} = \sum_{i=1}^{p} \beta_i X_{t+p-i} + \sigma_t \eta_t, \tag{2.7}$$

where η_t is white noise and σ_t is the seasonal volatility. The order p of the appropriate $AR(p)$ model is suggested by plotting the Partial Autocorrelation Function (PACF) of X_t and confirmed by the Bayesian Information Criterion, see Hurvich and Tsai (1989), for each city.

After seasonality and trend were removed a clear periodic pattern of the obtained residuals is shown which reveals that (2.7) is a heteroscedastic process. In order to model the seasonal variation of the residuals, we estimate the seasonal variance σ_t^2 of the residuals. We first divide the $AR(p)$ residuals into 365 groups, so that each group corresponds to the same day of the year and calculate the average of squared residuals for each group. Then the seasonal variation of residuals is calibrated with four different models.
Benth et al. (2007) proposed the one step model of Fourier terms:

$$\hat{\sigma}^2_{t,FTS} = c_1 + \sum_{i=1}^{L} \left\{ c_{2i} \cos\left(\frac{2i\pi t}{365}\right) + c_{2i+1} \sin\left(\frac{2i\pi t}{365}\right) \right\} \tag{2.8}$$

and found that L=4 gives a very good fit to estimate the variance function σ_t^2.
In order to constrain the number of parameters of equation (2.8), Härdle et al. (2011a) proposed to smooth the data with a Local Linear Regression $\hat{\sigma}^2_{t,LLR}$ estimator:

$$\hat{\sigma}^2_{t,LLR} = \arg\min_{a,b} \sum_{i=1}^{365} \left\{ \tilde{\sigma}_t^2 - a(t) - b(t)(t - t_0) \right\}^2 K\left(\frac{t - t_0}{h}\right) \tag{2.9}$$

where $\tilde{\sigma}_t^2$ is the average of squared residuals on each day t over all years, $t = 1, \ldots, 365$, h is a bandwidth and $K(\cdot)$ is a Kernel.
Alternatively a two steps model in Campbell and Diebold (2005) is considered. They suggest an additive model of a truncated Fourier function and a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) of the form

$$\hat{\sigma}^2_{t,FTSG} = c_1 + \sum_{i=1}^{L} \left\{ c_{2i} \cos\left(\frac{2i\pi t}{365}\right) + c_{2i+1} \sin\left(\frac{2i\pi t}{365}\right) \right\} + \alpha_1 \varepsilon^2_{t-1} + \beta_1 \sigma^2_{t-1}. \tag{2.10}$$

The Fourier series captures the seasonal volatility and the GARCH process captures the remaining volatility persistence effect of deseasonalized and detrended daily average temperatures.
Instead of considering an additive model for the calculation of the variance, Benth and Saltyte

Benth (2011) suggested a multiplicative model of Fourier and GARCH terms:

$$\hat{\sigma}^2_{t,MFTSG} = \left(c_1 + \sum_{i=1}^{L} \left\{ c_{2i} \cos\left(\frac{2i\pi t}{365}\right) + c_{2i+1} \sin\left(\frac{2i\pi t}{365}\right) \right\} \right) * \left(\alpha_1 \varepsilon^2_{t-1} + \beta_1 \sigma^2_{t-1}\right). \quad (2.11)$$

This multiplicative approach was motivated by the fact that GARCH effects were found after the calibration with $\hat{\sigma}_{t,FTS}$ as $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTS}}$, where $\hat{\varepsilon}_t$ are the residuals estimated by the $AR(p)$ process and due to the positivity of the variance.

We found out that $L = 4$ gives a good fit for estimating the seasonal variance of models (2.8), (2.10) and (2.11). Next, the residuals $\hat{\varepsilon}_t$ of $AR(p)$ are standardized by the seasonal volatilities which are estimated from the previous proposed models and are tested for normality. In order to check normality, two different tests are applied with null hypothesis that the data follow a normal distribution:

- the Anderson-Darling test, defined as

  $A^2 = -n - S$, with $S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[lnF\left(Y_i\right) + ln\left(1 - F\left(Y_{n+1-i}\right)\right)\right]$,

  where F denotes the normal cumulative distribution function and Y_i are the ordered data
- and the Jarque-Bera test, constructed as

  $JB = \frac{n}{6}\left(S^2 + \frac{1}{4}\left(K-3\right)^2\right)$,

  where S is the skewness and K the kurtosis of the tested random variable.

3 Analysis of Temperature Dynamics in London, Rome, Paris and Amsterdam

The model of temperature dynamics is implemented on four european cities London, Rome, Paris and Amsterdam. We are interested on estimating the parameters of temperature dynamics on these cities since they cover a part of the industrial Blue Banana european area. The territory of Blue Banana covers one of the highest concentrations of people, money and industry in the world and the aim of our analysis is to detect whether there is an evidence of global warming in this specific indusrial area of Europe. The model of temperature dynamics (2.1) was also applied on other cities of Europe such as Stockholm, Sweden, see Benth et al. (2007) and Berlin, Germany, see Härdle and López Cabrera (2011).

Figure 3.1: Location of London, Rome, Paris and Amsterdam on the map of Europe. eumap.R

3.1 The Temperature Data

The temperature data for the four european cities were obtained from Bloomberg. The datasets are daily average temperatures T_t, measured in °C and are defined as $T_t = \frac{T_{t,max}+T_{t,min}}{2}$. The observations are from January 1, 1973 to October 10, 2009. There were 0.03% missing values in the data which were handled by computing the average of the previous and the next day's temperature. The leap years were removed in order the estimation procedure to be more consistent.

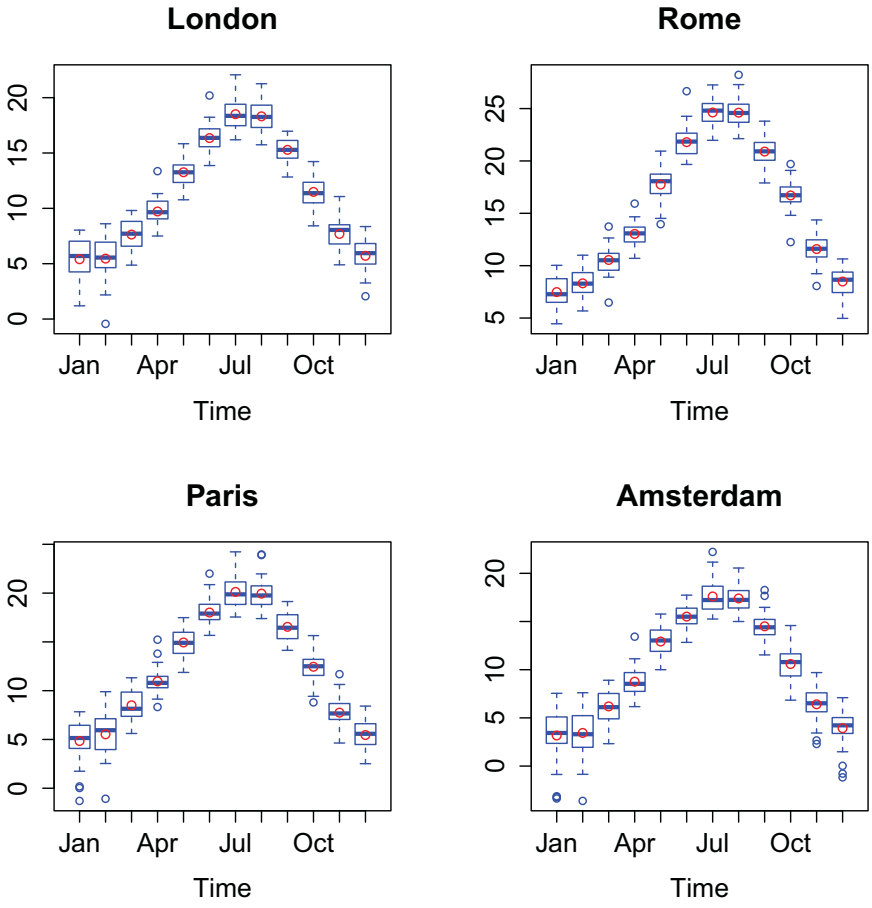

Figure 3.2: Monthly average temperatures and monthly temperature means (red circles) for London, Rome, Paris, Amsterdam from 01.01.1973 - 31.12.2008. boxplots.R

The data for each city is splitted in two datasets. The first data set consists of 13140 observations from January 1, 1973 to December 31, 2008 (in-sample data) which is used for model estimation. The second data set consists of 283 observations from January 1, 2009 to October 10, 2009 (out-of-sample data) and is used for model validation.

3.2 Seasonal Correction

We continue our analysis with the estimation of the seasonal components of the temperature data. Two methods are applied the Local Linear Regression (LLR) estimator and the Lasso regression. LLR given in (2.2), is fitted to the data using the Epanechnikov Kernel and a bandwidth proposed by Bowman and Azzalini (1997) and recently applied to the temperature data by Härdle et al. (2011a). The fitted seasonal component derived by LLR method is shown in Figure 3.3.

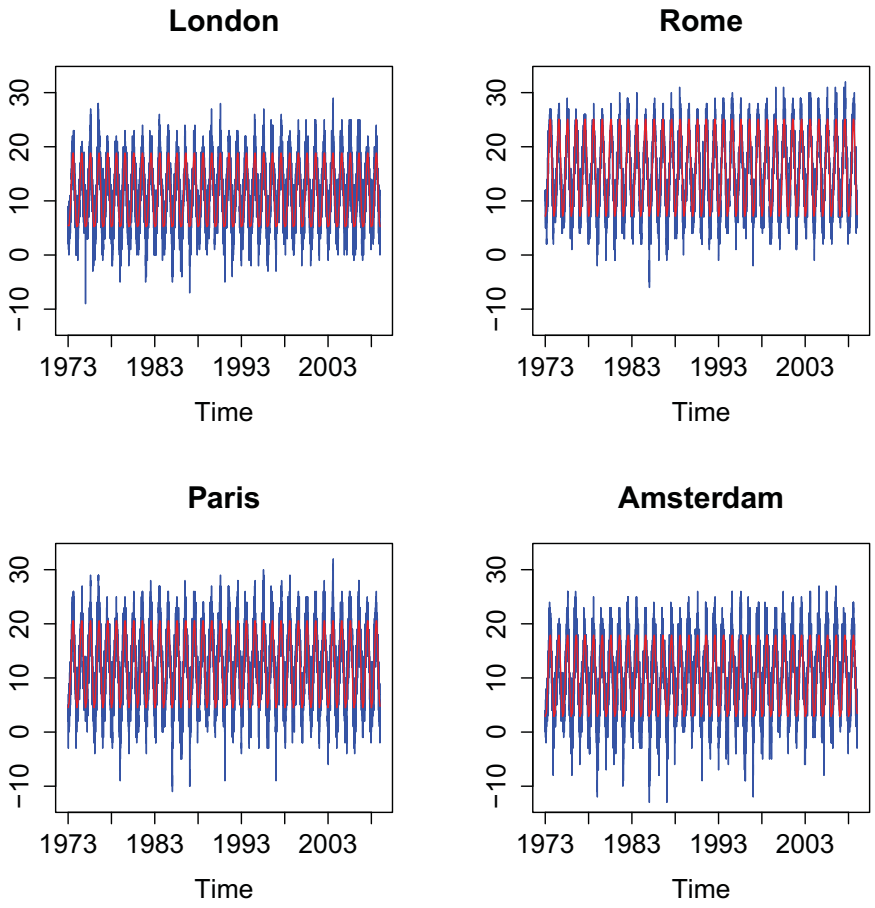

Figure 3.3: Daily average temperatures (blue line) and seasonality effect (red line), estimated with LLR, for the four european cities. seasonalityLLR.R

The Lasso regression is defined in formula (2.4). Our initial time basis (of size 43) follows Song et al. (2010) approach. The time basis is represented in Table 3.1. The orthogonal Legendre polynomial basis $u_1(t) = 1$, $u_2(t) = t$, $u_3(t) = 3t^2 - 1$ captures the global trend in time. The periodic variations, namely the seasonal and the large period effects, can be captured by the Fourier series: $u_4(t) = sin(2\pi t/p)$, $u_5(t) = cos(2\pi t/p)$, $u_6(t) = sin(2\pi t/(p/2))$, $u_7(t) = cos(2\pi t/(p/2))$,..., where $p = 365$ or $p = 365 \cdot 11$. The period of 11 years indicates the 11-year solar activity cycle, according to meteorologists.

	Factors		Factors
Trend	1	Large	$sin2\pi t/(365 \cdot 11)$
(Year by Year)	t	Period	$cos2\pi t/(365 \cdot 11)$
	$3t^2 - 1$		$sin4\pi t/(365 \cdot 11)$
Seasonal	$sin2\pi t/365$		$cos4\pi t/(365 \cdot 11)$
Effect	$cos2\pi t/365$		$sin6\pi t/(365 \cdot 11)$
	...		...
	$cos20\pi t/365$		$sin20\pi t/(365 \cdot 11)$

Table 3.1: Initial choice of time basis.

Lasso minimizes the sum of squared errors and penalizes the coefficients. The selection of the coefficients for the estimation of the seasonality function of equation (2.3) is done by Lasso algorithm, see equation (2.4). From the broad class of initial basis functions just few are selected. Therefore, we avoid overfitting and just significant terms are incorporated. Thus a better fit for the seasonality is achieved. Figure 3.4 shows how the penalty term λ shrinks the coefficients β of equation (2.3). The different colors correspond to the time basis of Table 3.1 which is taken into account in order to select the coefficients.

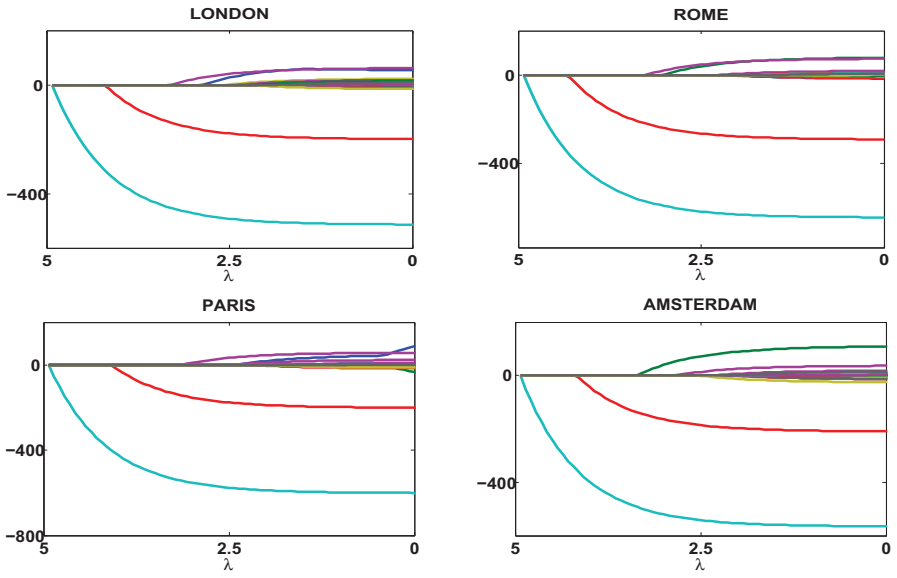

Figure 3.4: Shrinkage of the coefficients to zero via the Lasso penalty for the four european cities. seasonalityLasso.m

In order to avoid overspecification and improve the selection of seasonal component (order of time trends and fourier series) the Lasso regression was applied for the sequel of our analysis. According to the time basis which is selected after the Lasso fit the seasonality models for the four european capitals are the following:

Parameters	London	Rome	Paris	Amsterdam
α	10.752	15.019	11.797	9.389
β	$8 \cdot 10^{-5}$	-	$5 \cdot 10^{-5}$	-
γ	-	$3 \cdot 10^{-9}$	-	$4 \cdot 10^{-9}$
δ_1	6.806	8.725	7.826	7.444
ζ_1	204.307	207.161	201.241	203.146
δ_2	-8.208	-0.933	-7.222	-0.460
ζ_2	126.571	131.614	136.806	136.171
δ_5	243.601	0.244	0.296	-
ζ_5	1.498	146.65	-1095.272	-
κ_5	-	-	-	-0.366
ξ_5	-	-	-	305.919

Table 3.2: Estimated parameters with nonlinear least squares of the seasonality models w.r.t. the time basis selected by Lasso. seasonalityLasso.R

London:

$$\Lambda_{t,L} = \alpha + \beta t + \delta_1 cos\{2\pi(t-\zeta_1)/365\} + \delta_2 cos\{4\pi(t-\zeta_2)/365\} + \delta_5 cos\{10\pi(t-\zeta_5)/365\} \tag{3.1}$$

Rome:

$$\begin{aligned} \Lambda_{t,R} &= \alpha + \gamma(3t^2-1) + \delta_1 cos\{2\pi(t-\zeta_1)/365\} + \delta_2 cos\{4\pi(t-\zeta_2)/365\} \\ &+ \delta_5 cos\{10\pi(t-\zeta_5)/365\} \end{aligned} \tag{3.2}$$

Paris:

$$\Lambda_{t,P} = \alpha + \beta t + \delta_1 cos\{2\pi(t-\zeta_1)/365\} + \delta_2 cos\{4\pi(t-\zeta_2)/365\} + \delta_5 cos\{10\pi(t-\zeta_5)/365\} \tag{3.3}$$

Amsterdam:

$$\begin{aligned} \Lambda_{t,A} &= \alpha + \gamma(3t^2-1) + \delta_1 cos\{2\pi(t-\zeta_1)/365\} + \delta_2 cos\{4\pi(t-\zeta_2)/365\} \\ &+ \kappa_5 cos\{10\pi(t-\xi_5)/(365 \cdot 11)\} \end{aligned} \tag{3.4}$$

The parameters of the seasonality models of London, Rome, Paris and Amsterdam are estimated with the non-linear least squares algorithm and are displayed in Table 3.2. The estimated parameter α stands for the average of the temperature which is higher for Rome as it was expected. The coefficients β and γ of linear and quadratic terms represent the global warming effect. Parameters δ_1, δ_2, δ_5 and κ_5 are the maximum displacements of the periodic terms (cosine functions) while ζ_1, ζ_2, ζ_5 and ξ_5 are their shifts. Since most of the coefficients are positive, we can claim that there is an evidence for global warming. Our conclusion is also confimed by Figure 3.5 which shows an upward trend of the temperature for all cities, particularly for Rome and Amsterdam.

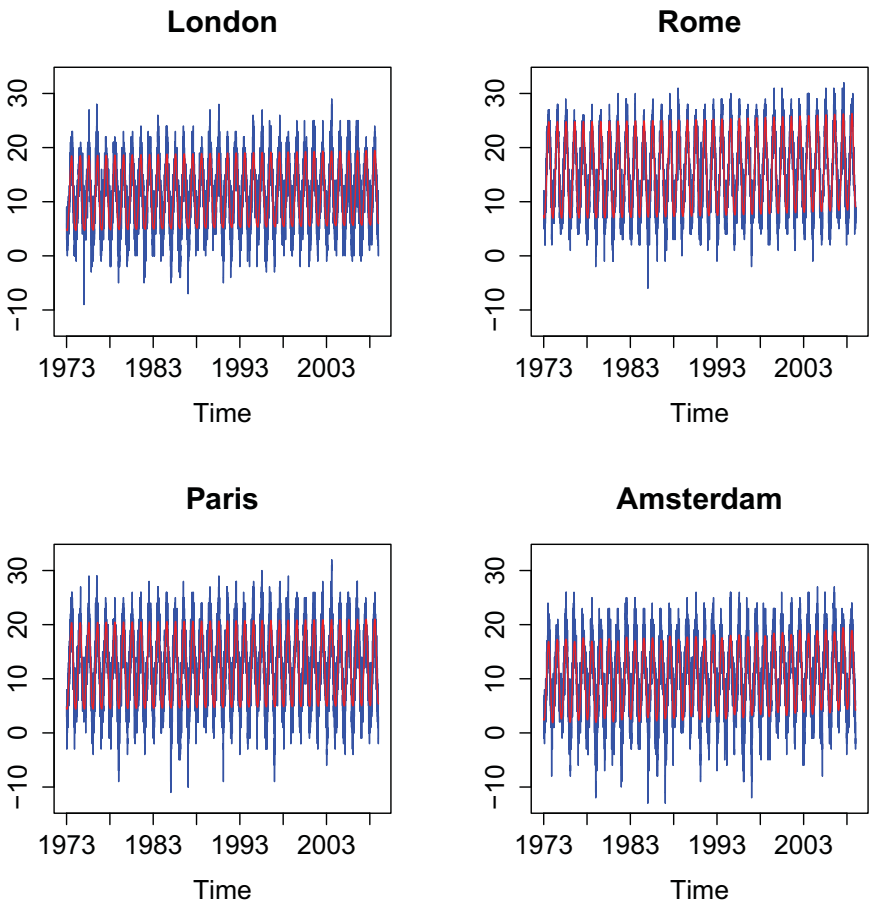

Figure 3.5: Daily average temperatures (blue line) and seasonality effect (red line), estimated with Lasso, for the four european cities. seasonalityLasso.R

3.3 Trend Correction

We remove the seasonality from the daily average series and plot the residuals (Figure 3.6):

$$X_t = T_t - \Lambda_t. \tag{3.5}$$

Figure 3.6 suggests that an autoregressive process should be fitted to the residual series. Before we fit the AR process we check whether the residuals are stationary with ADF test (2.5) and KPSS (2.6) test. According to Table 3.3, the ADF test rejects the H_0 of unit root and the KPSS test can not reject the H_0 of stationarity. Thus, an AR process can be fitted to (3.5) in order to model the temperature evolution of each city. The partial autocorrelation functions (PACF) of Figure 3.7, show that higher order autoregressive processes should be fitted to the residuals. The order p of the $AR(p)$ models is selected by the Bayesian Information Criterion (BIC). In Figure 3.8, we can see that BIC is maximized for order $p = 4$ for London and for order $p = 3$ for Rome, Paris and Amsterdam.

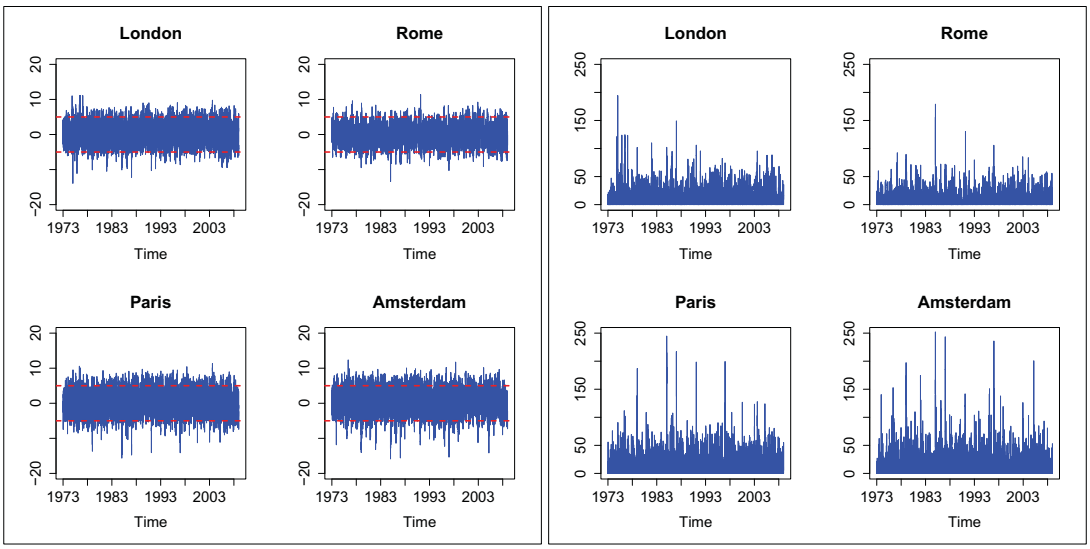

Figure 3.6: Residuals (left panel) and squared residuals (right panel) after removing the seasonality. seasonalityLasso.R

City	$\hat{\tau}$(p-value)	$\hat{k}$(p-value)
London	-20.29(<0.01)	0.167(<0.1)
Rome	-18.67(<0.01)	0.094(<0.1)
Paris	-20.66(<0.01)	0.221(<0.1)
Amsterdam	-20.05(<0.01)	0.070(<0.1)

Table 3.3: ADF and KPSS stationarity tests. autoregressive.R

Next, the AR processes with orders according to the BIC, are fitted to the residuals and the coefficients of the processes are estimated, Table 3.4.

After deseasonalizing and detrending the daily temperature time series, the residuals $\hat{\varepsilon}_t$ and squared residuals $\hat{\varepsilon}_t^2$ of equation (2.7) are plotted in Figure 3.9. Additionally, the autocorrelation functions (ACF) are calculated and presented in Figure 3.10. It is depicted a clear seasonal pattern of squared residuals which motivates us to estimate the average seasonal variance of the $AR(p)$ residuals for each day over all years and then calibrate the seasonal variation with the four models presented in equations (2.8), (2.9), (2.10) and (2.11). Figures 3.11 and 3.12 show the average of the squared residuals over 36 years (from 1973 to 2008) for each day of the year and the estimated squared volatility functions $\hat{\sigma}_{t,LLR}$, $\hat{\sigma}_{t,FTS}$, $\hat{\sigma}_{t,FTSG}$ and $\hat{\sigma}_{t,MFTSG}$. We

Coefficients	London	Rome	Paris	Amsterdam
β_1	0.759	0.818	0.909	0.888
β_2	-0.070	-0.085	-0.194	-0.187
β_3	0.016	0.033	0.065	0.084
β_4	0.036	-	-	-

Table 3.4: Coefficients of AR(p), Model selection: BIC. autoregressive.R

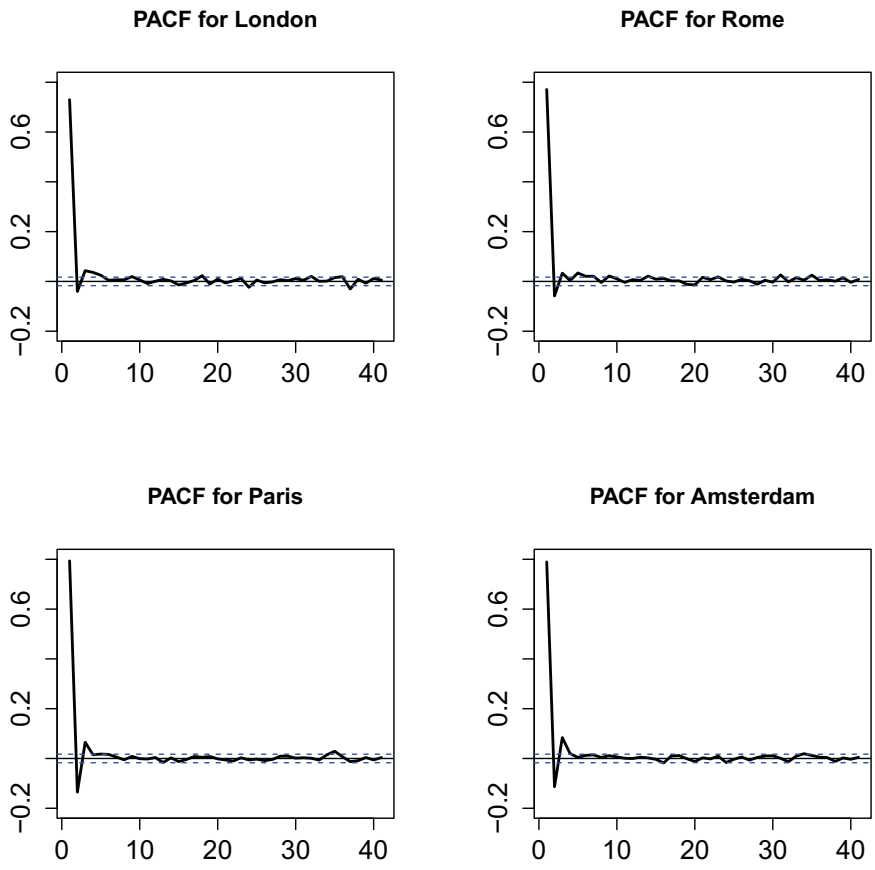

Figure 3.7: Partial autocorrelation funtion (PACF) for the four european cities after the seasonal correction. autoregressive.R

conclude that for London and Rome the variance is higher in winter and lower during summer. For Paris and Amsterdam the variance has similar structure, higher values in early winter and late spring and lower values in late summer and early spring.

Next, the residuals $\hat{\varepsilon}_t$ are standardized by the estimated volatility of the four models. Figure 3.14 shows the ACF of the standardized residuals and squared standardized residuals for Rome and we conclude that no seasonal pattern exists anymore. The standardized residuals are checked for normality with Anderson-Darling and Jarque-Bera tests so that we can decide which estimated seasonal squared volatility function is the best fit. Table 3.5 shows the skewness, kurtosis and tests statistics for the standardized residuals ($\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTS}}$, $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,LLR}}$, $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTSG}}$ and $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,MFTSG}}$). We conclude that the residuals divided with a multiplicative structure of the seasonality and the GARCH effect in the seasonal variation seems to be closer to normal. This is also confirmed with the QQ-plots shown in Figure 3.13.

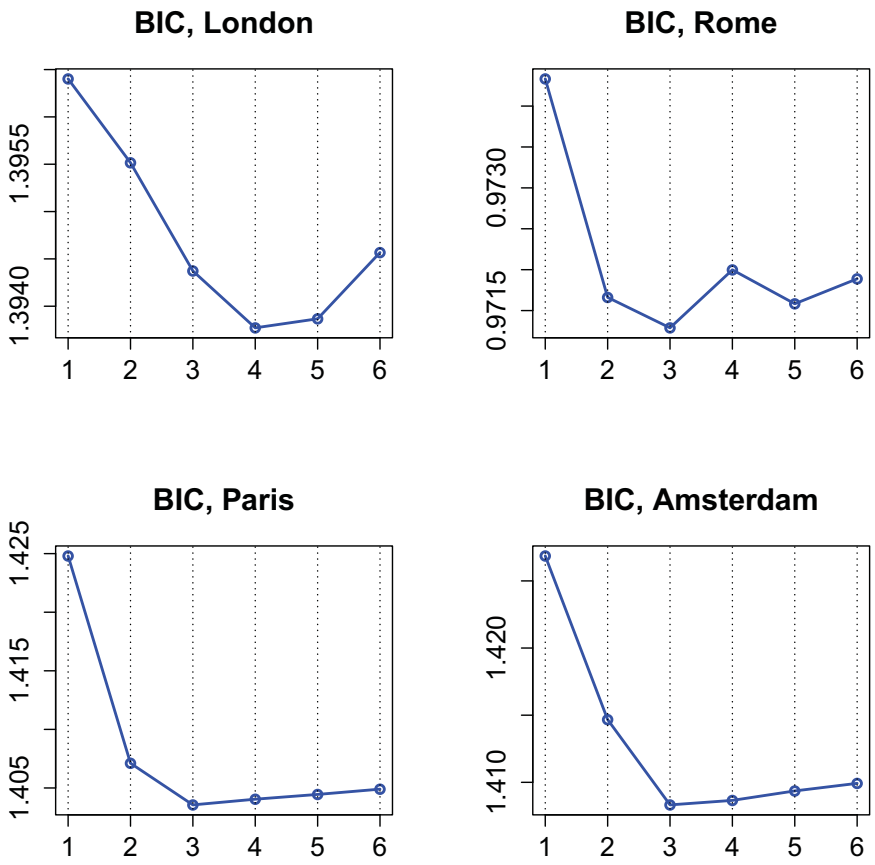

Figure 3.8: Bayesian Information Criterion (BIC) for the four european cities after the seasonal correction. autoregressive.R

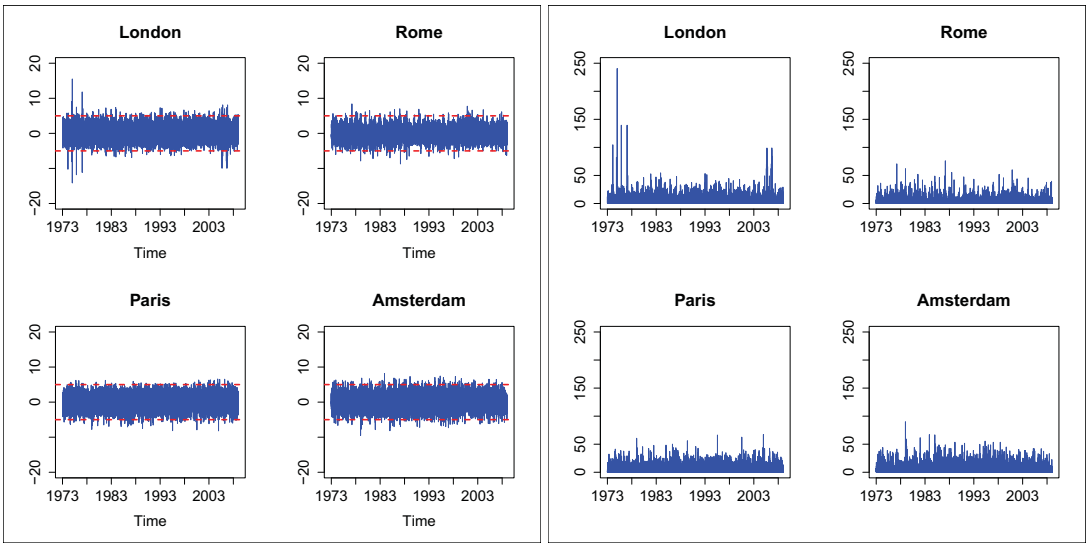

Figure 3.9: Residuals (left panel) and squared residuals (right panel) after removing the seasonality and trend. autoregressive.R

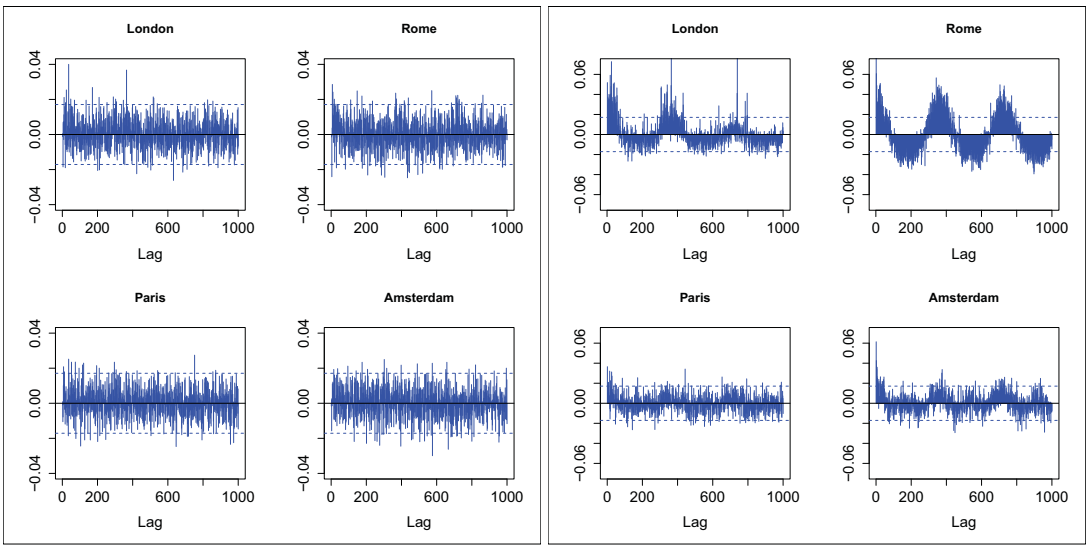

Figure 3.10: ACF of residuals (left panel) and squared residuals (right panel) after removing the seasonality and trend. autoregressive.R

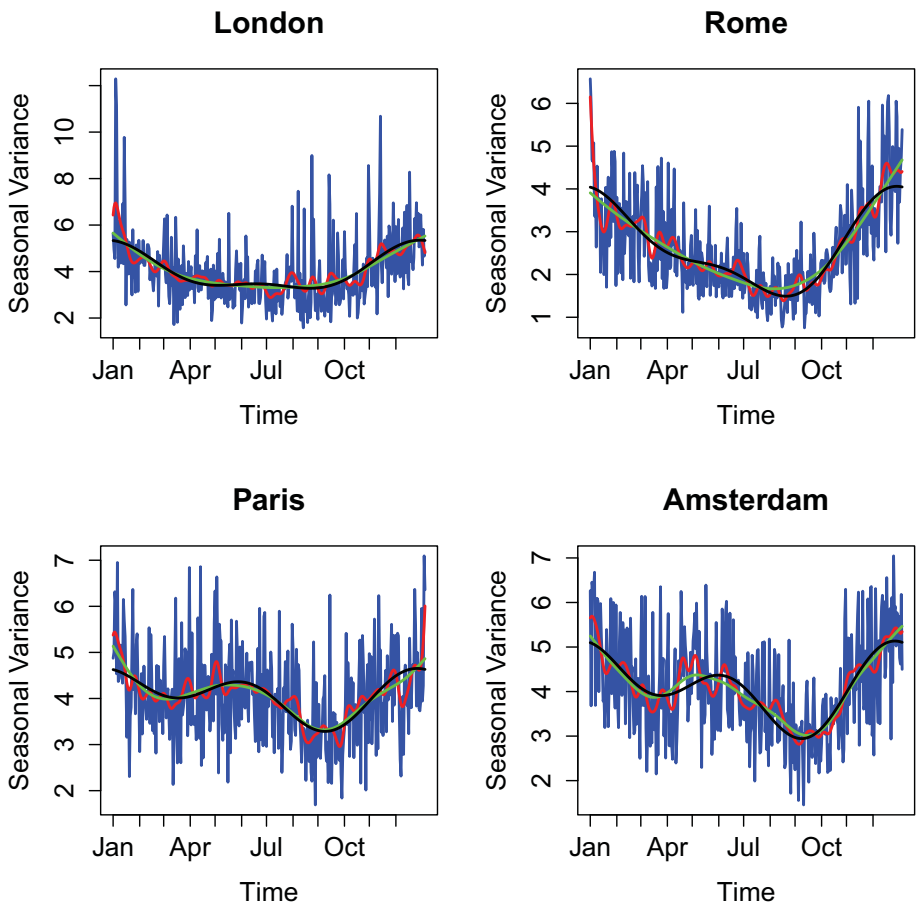

Figure 3.11: Estimated seasonal variance (blue line), $\hat{\sigma}_{t,LLR}$ (red line), $\hat{\sigma}_{t,FTS}$ (black line) and $\hat{\sigma}_{t,FTSG}$ (green line) for each city. seasonalCalibration.R

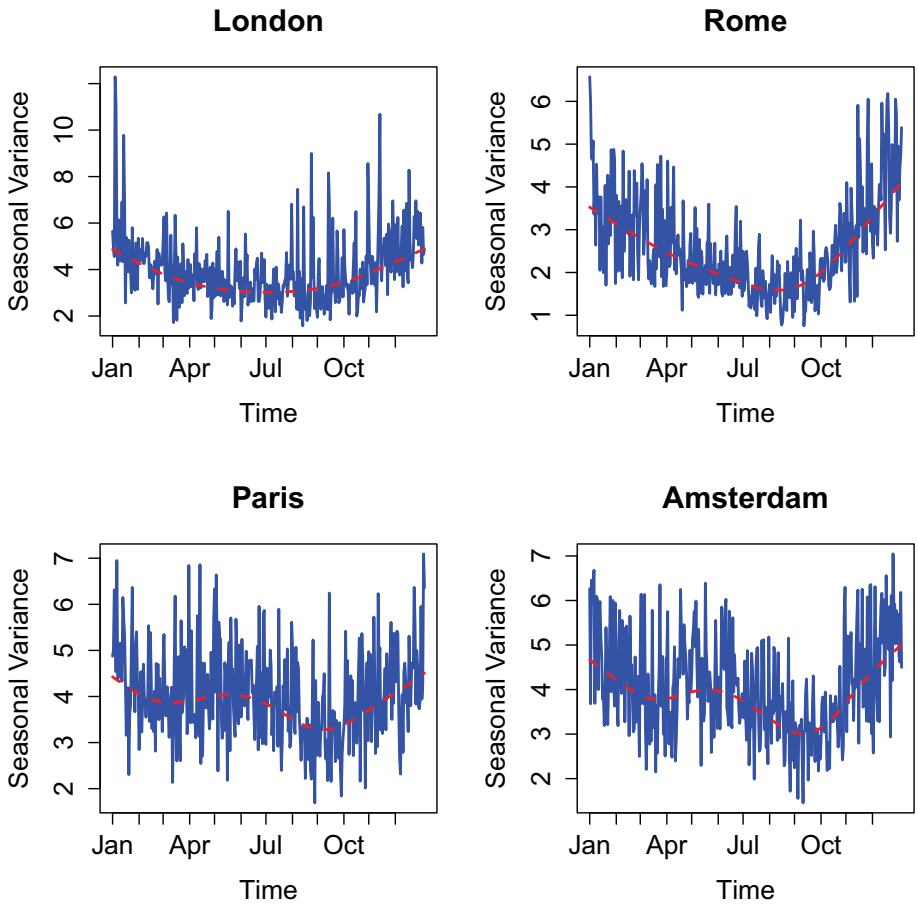

Figure 3.12: Estimated seasonal variance (blue line) and $\hat{\sigma}_{t,MFTSG}$ (red dashed line) for each city. seasonalCalibration.R

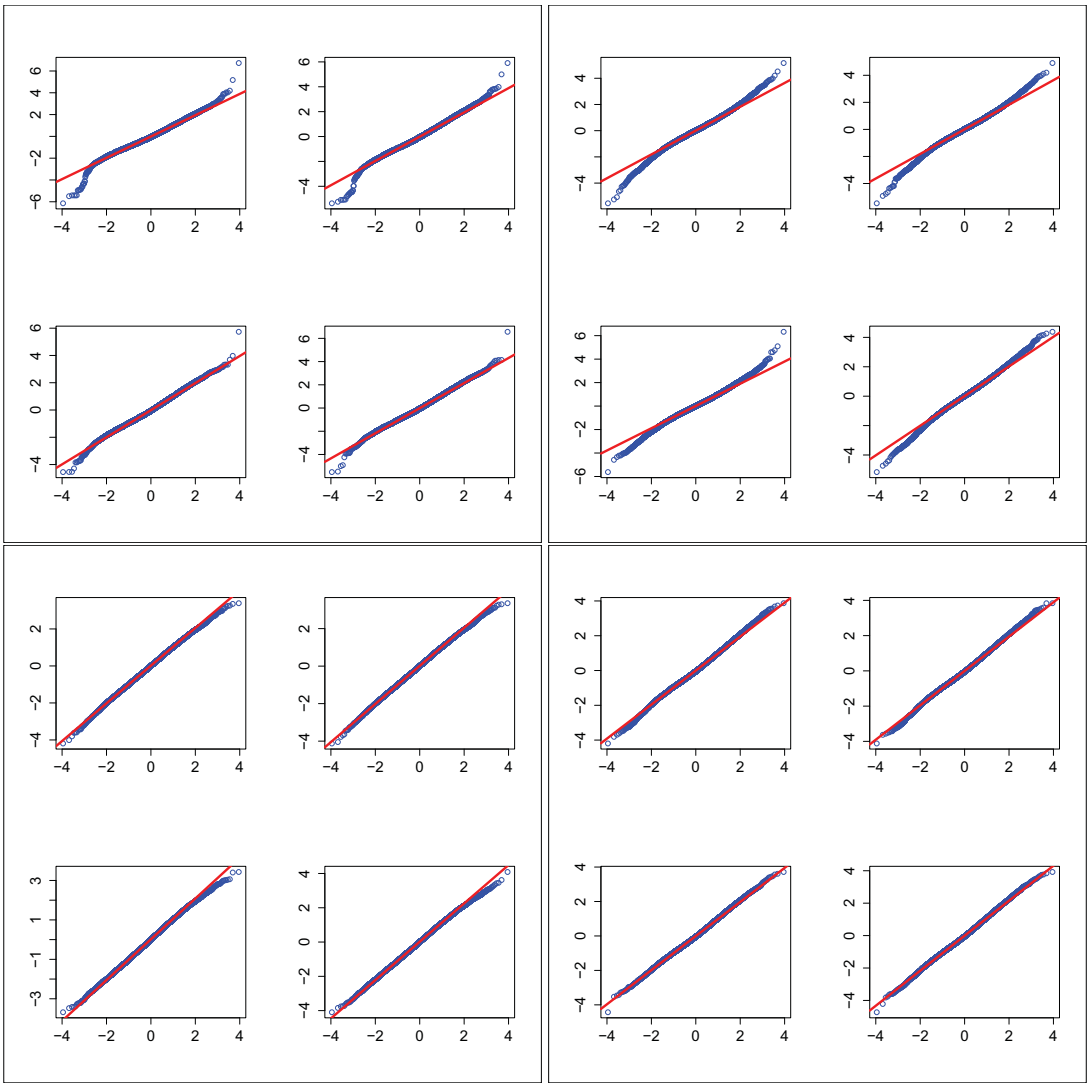

Figure 3.13: QQ-plots of standardized residuals $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTS}}$ (upper left), $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,LLR}}$ (upper right), $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTSG}}$ (lower left) and $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,MTSG}}$ (lower right) for each panel, for London (upper left panel), Rome (upper right panel), Paris (lower left panel) and Amsterdam (lower right panel). seasonalCalibration.R

City		$\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTS}}$	$\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,LLR}}$	$\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTSG}}$	$\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,MFTSG}}$
London	Anderson Darling	13.724	13.268	11.527	10.649
	Jarque Bera	593.779	360.609	61.531	78.338
	Kurtosis	4.034	3.797	3.216	3.272
	Skewness	0.062	0.076	0.128	0.131
Rome	Anderson Darling	18.382	16.482	14.449	10.770
	Jarque Bera	615.057	509.693	405.052	197.498
	Kurtosis	4.036	3.943	3.853	3.571
	Skewness	-0.113	-0.102	-0.054	-0.094
Paris	Anderson Darling	0.952	1.010	1.975	1.703
	Jarque Bera	13.121	12.237	26.768	22.339
	Kurtosis	2.960	2.933	2.797	2.832
	Skewness	-0.074	-0.067	-0.044	-0.056
Amsterdam	Anderson Darling	9.354	9.015	7.270	7.032
	Jarque Bera	57.701	50.169	25.782	23.804
	Kurtosis	3.253	3.221	3.051	3.088
	Skewness	0.102	0.103	0.105	0.095

Table 3.5: Anderson-Darling and Jarque Bera normality tests as well as skewness and kurtosis for the standardized residuals. seasonalCalibration.R

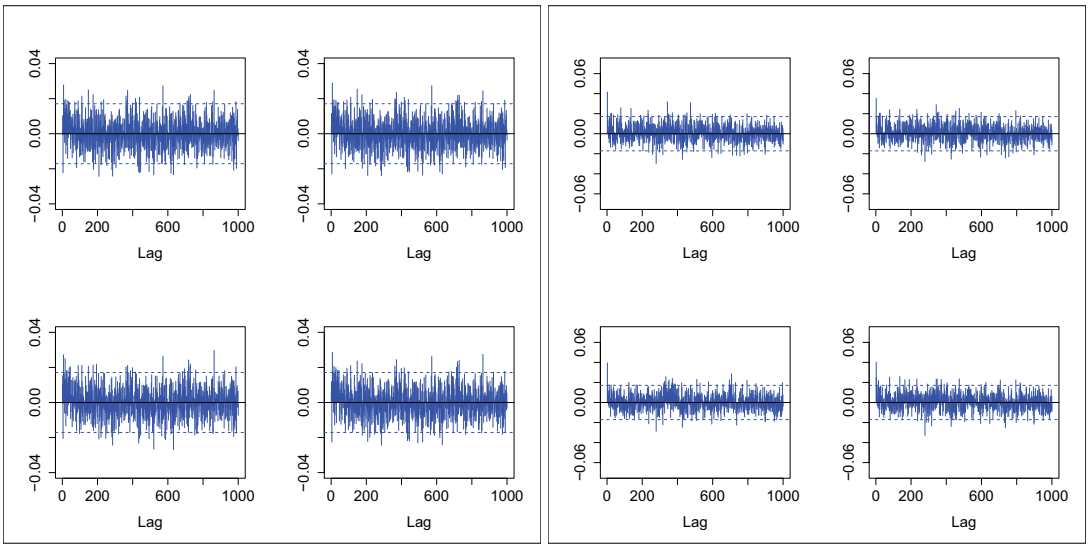

Figure 3.14: ACF of residuals (left panel) and squared residuals (right panel), $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTS}}$ (upper left), $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,LLR}}$ (upper right), $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,MFTSG}}$ (lower left), $\frac{\hat{\varepsilon}_t}{\hat{\sigma}_{t,FTSG}}$ (lower right) for each panel for Rome. seasonalCalibration.R

4 Model Validation

Data from January 1, 2009 to October 10, 2009 of 283 observations are available for each city and are used for model validation. We generate one step ahead predictions from the 283 out of sample observations with respect to the multiplicative model of Fourier and GARCH terms. The observed and predicted values are shown in Figure 4.1. The differences between the two lines (red color-real values and blue color-forecasts) correspond to the prediction errors (PE). The green lines correspond to the 95% pointwise confidence intervals.

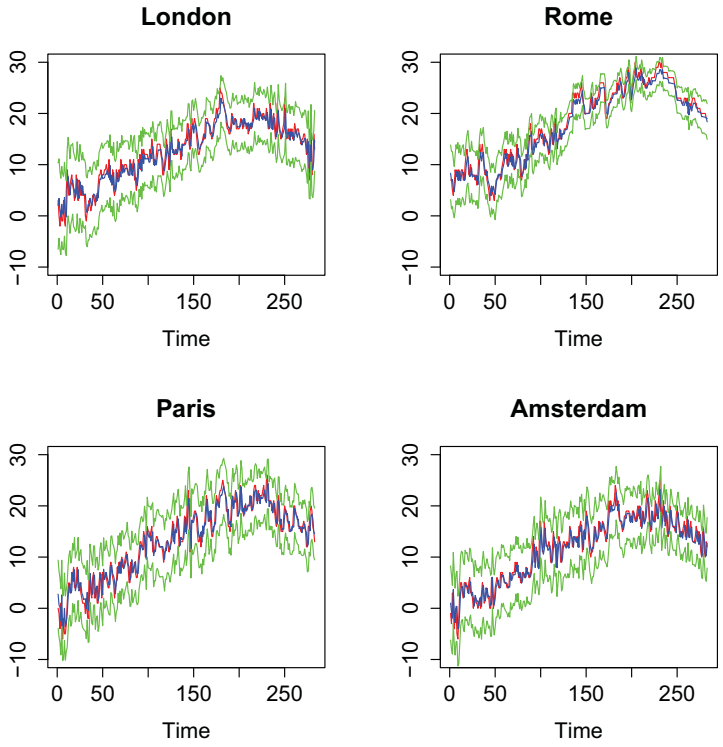

Figure 4.1: Observed (red line), predicted (blue line) values and prediction intervals (green line).
modelValidation.R

Jarque-Bera tests are applied for checking whether the PEs follow a normal distribution. The normality of the PEs is not rejected at 5% significance level for all analyzed cities, the results are shown in Table 4.1. The kurtosis and skewness of PEs are reported in Table 4.2. The PEs' skewness of London and Amsterdam is greater than 0. Therefore prediction values derived by the fitted model are more often below the observed temperature. For Rome the skewness is negative, the forecasted temperatures are more often above the real observations. For Paris we report PEs' skewness is close to 0. The kurtosis of the PEs' distributions are reported to be

City	*JB*	(p-value)
London	44.733	(< 0.001)
Rome	27.326	(< 0.001)
Paris	7.176	(0.028)
Amsterdam	12.576	(0.002)

Table 4.1: Jarque-Bera Tests for prediction errors. modelValidation.R

City	*Kurtosis*	*Skewness*
London	4.889	0.235
Rome	4.075	-0.539
Paris	3.769	-0.067
Amsterdam	3.945	0.208

Table 4.2: Kurtosis and skewness of prediction errors. modelValidation.R

leptokurtic. Moreover, QQ-plots of Figure 4.2 suggest that the PEs for the four cities are close to the normal distribution with Paris to satisfy the best approximation.

In order to test the out-of-sample forecasts two accuracy measures are applied, the root mean squared prediction error (RMSE) given by $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$ and the mean absolute error (MAE) defined as $MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i|$, see Hyndman and Koehler (2006). RMSE and MAE are shown in Table 4.3. The best forecasting performance is achieved for the city of Rome. The prediction power of the fitted model for the other cities is relatively similar. Additionally from Table 4.3 it is clear the RMSE and MAE have very small values. Therefore, we conclude that the multiplicative model gives us quite precise one day ahead predictions and it is a good model for forecasting.

Moreover, 95% and 80% predictions intervals (PI) were calculated from the model. Concerning the calculation of PIs we followed Benth and Saltyte Benth (2011) method. 283 random innovations were generated. Secondly, a series of values was built from the model. This iteration was repeated 1000 times, 1000 realisations of the model were simulated. The PIs were then computed as a corresponding pointwise (for all 283 data points) empirical quantile. In the next step we once more simulated 1000 trajectories and investigated the robustness of constructed PIs. Table 4.4 shows the percentages of the simulated observations which lie outside the constructed PIs.

City	*RMSE*	*MAE*
London	1.999	1.489
Rome	1.486	1.178
Paris	2.025	1.555
Amsterdam	1.934	1.475

Table 4.3: Forecast accuracy measures. modelValidation.R

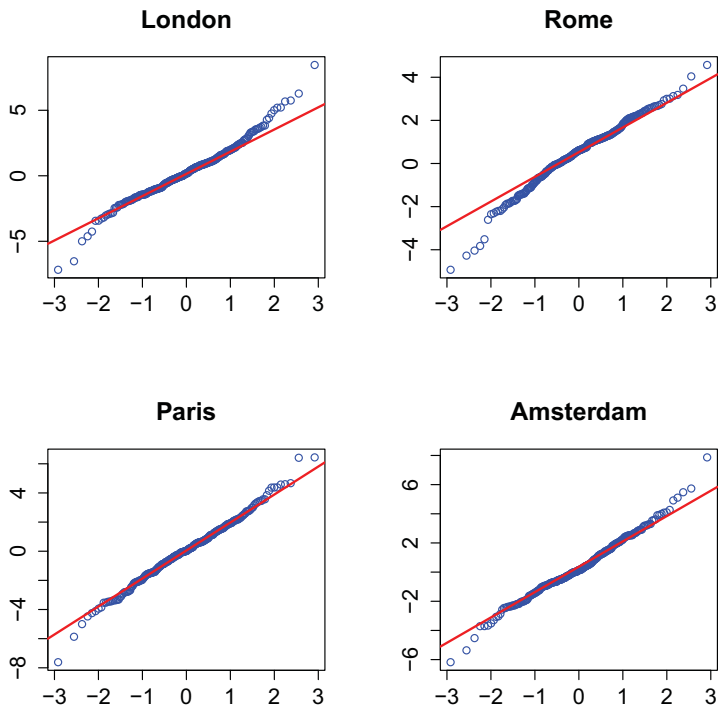

Figure 4.2: QQ-plots for prediction errors. modelValidation.R

City	95%	80%
London	5.069	20.076
Rome	5.063	20.008
Paris	5.052	20.055
Amsterdam	5.095	20.053

Table 4.4: Prediction intervals. modelValidation.R

5 Expectile Curves

The seasonal variation from the fitted temperature model of equation 2.7 can also be analyzed by expectile curves, as proposed by Guo and Härdle (2012). This approach allows us to investigate the extreme temperatures reported in the data - the special behavior of "non-average" observations. The τ-conditional expectile $v_\tau(x)$, $0 < \tau < 1$, given x, is defined as

$$v(x) = \arg\min_{\theta} \mathrm{E}\left\{\rho_\tau(y-\theta)|X = x\right\}, \tag{5.1}$$

where $\rho_\tau(u) = |\mathbf{1}(u \leq 0) - \tau|u^2$ is the loss function. Note that $\rho^*_\tau(u) = |\mathbf{1}(u \leq 0) - \tau|u$ leads to quantile regression framework, also discussed in this paper. It was shown, that there is one-to one mapping relationship between quantile and expectile, see Yao and Tong (1996). Guo and Härdle (2012) introduced the localized nonlinear smoother $v_n(x)$ of the expectile regression curve and constructed a confidence corridor around the estimated expectile function of the conditional distribution of Y given x.

Following the methodology of Guo and Härdle (2012) we apply the expectiles to the squared residuals for each day of the year over each 12 year period. The seasonal and AR effects were first removed. $X = 1, \ldots, 365$ denotes the day of the year and Y are the squared model residuals within each 12-year subsample.

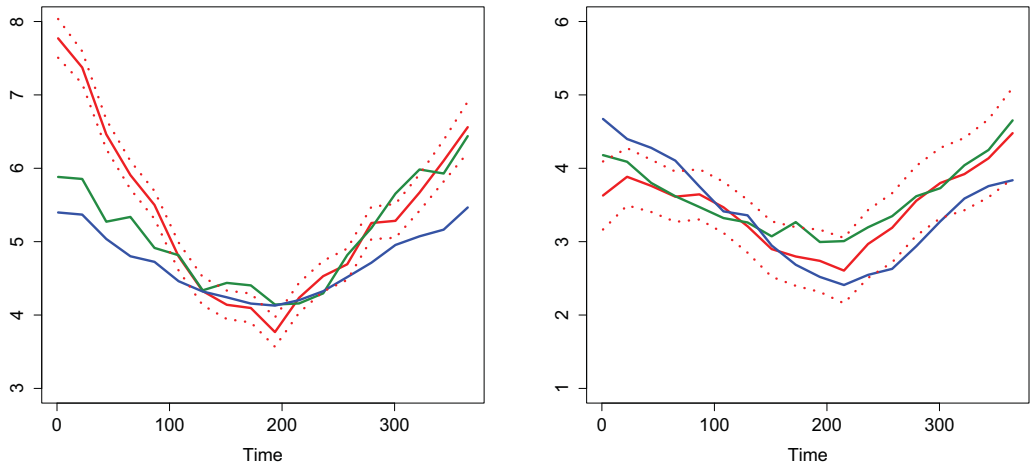

Figure 5.1: 0.9 - expectile curves for London (left) and Rome (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the first 12 years expectile. expectiles.R

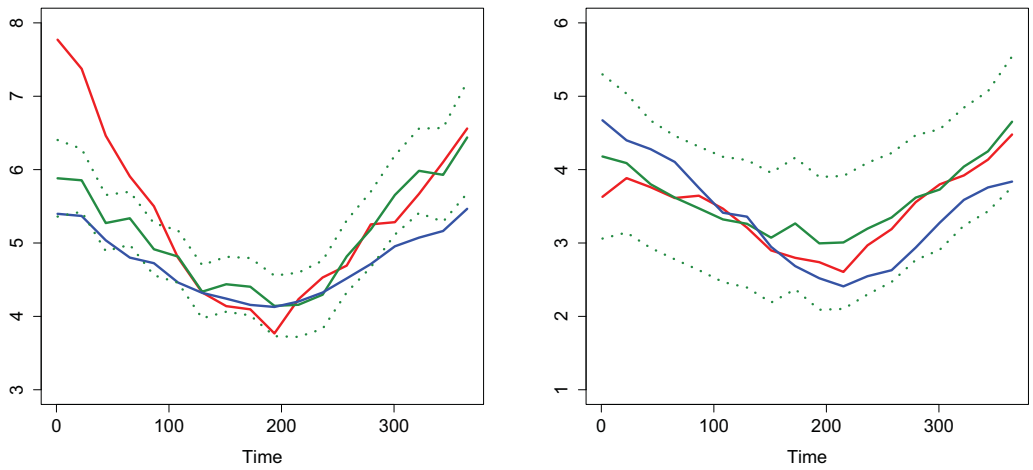

Figure 5.2: 0.9 - expectile curves for London (left) and Rome (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the second 12 years expectile. expectiles.R

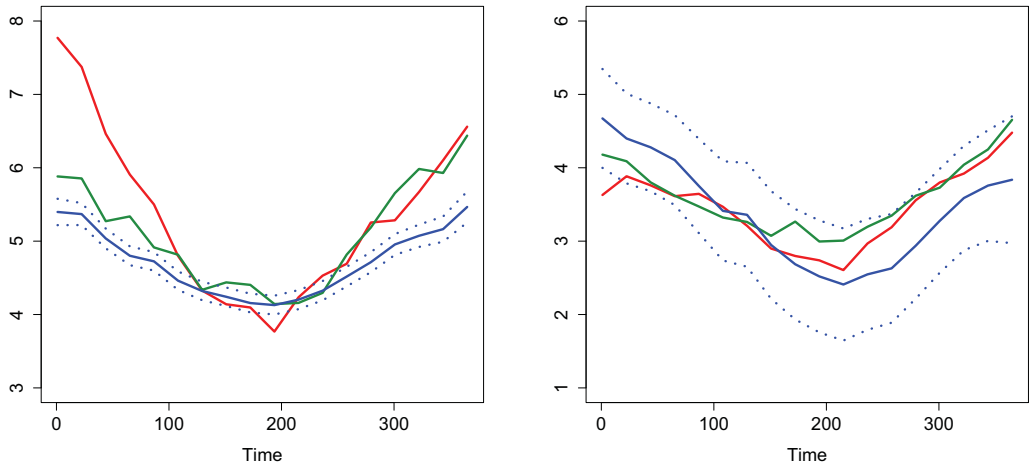

Figure 5.3: 0.9 - expectile curves for London (left) and Rome (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the latest 12 years expectile. expectiles.R

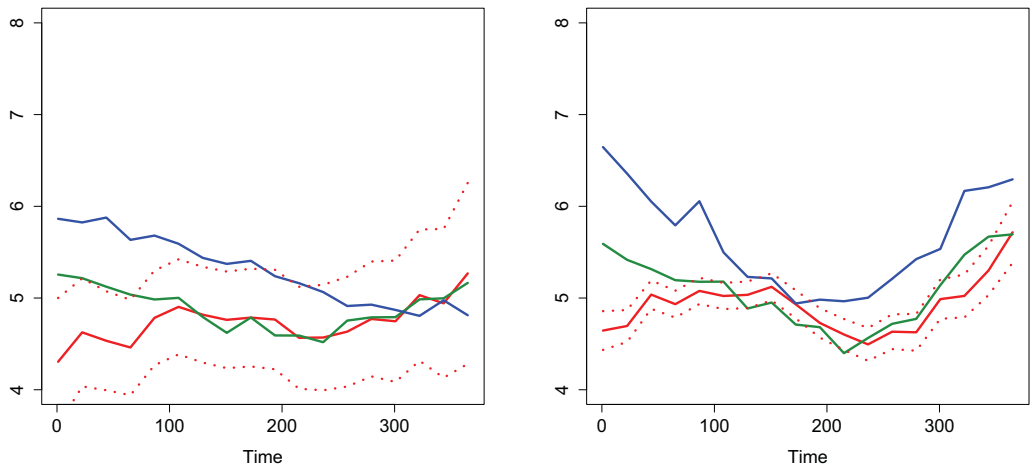

Figure 5.4: 0.9 - expectile curves for Paris (left) and Amsterdam (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line for the latest 12 years, with the 5% - 95% confidence corridors for the first 12 years expectile. expectiles.R

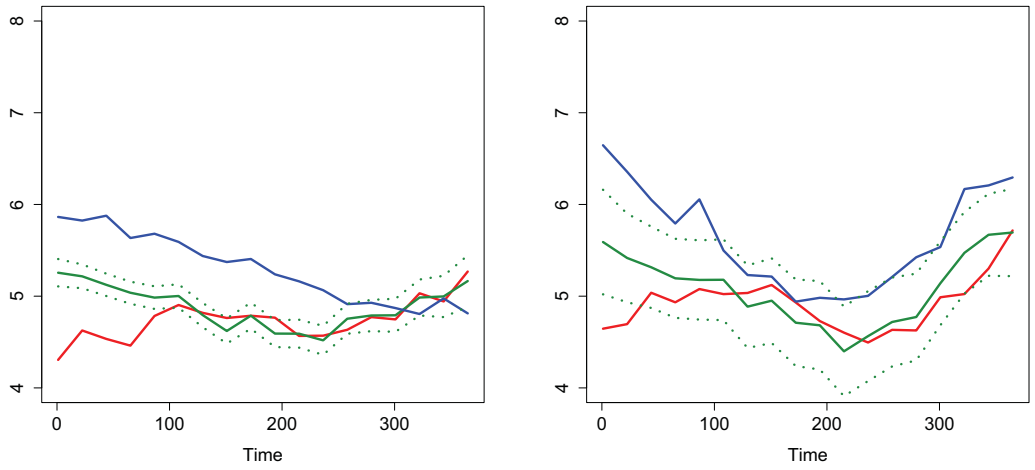

Figure 5.5: 0.9 - expectile curves for Paris (left) and Amsterdam (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line for the latest 12 years, with the 5% - 95% confidence corridors for the second 12 years expectile. expectiles.R

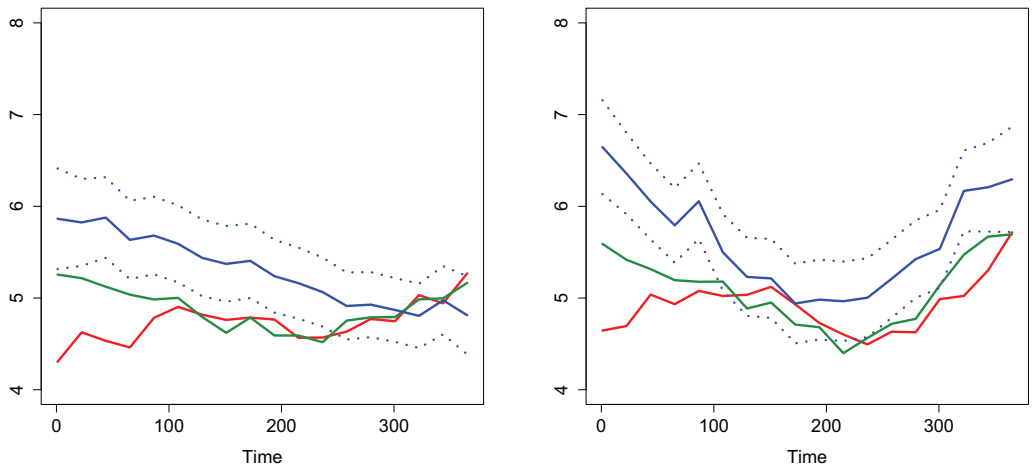

Figure 5.6: 0.9 - expectile curves for Paris (left) and Amsterdam (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line for the latest 12 years, with the 5% - 95% confidence corridors for the latest 12 years expectile. expectiles.R

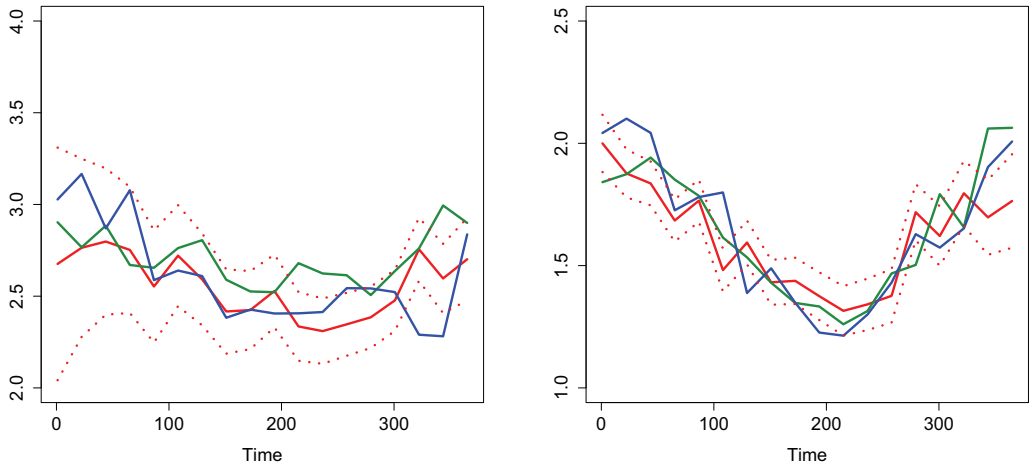

Figure 5.7: 0.01 - expectile curves for London (left) and Rome (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the first 12 years expectile. expectiles.R

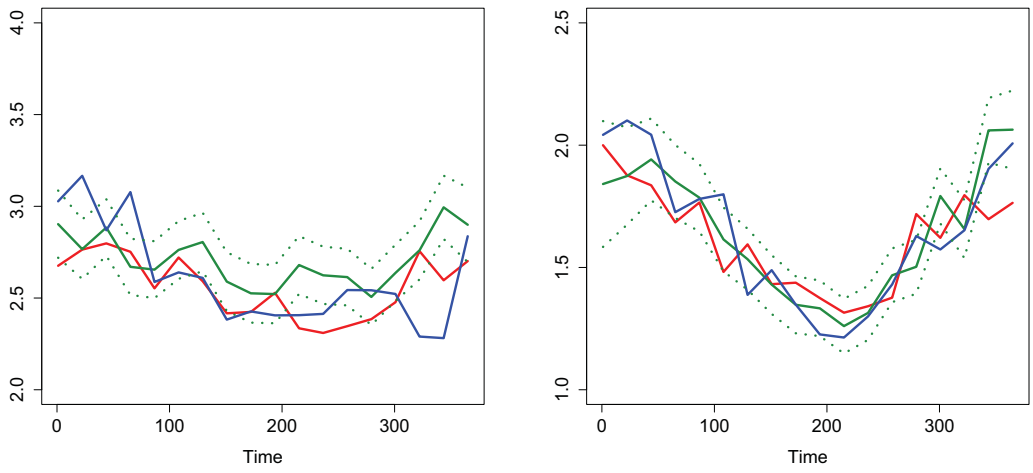

Figure 5.8: 0.01 - expectile curves for London (left) and Rome (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the second 12 years expectile. expectiles.R

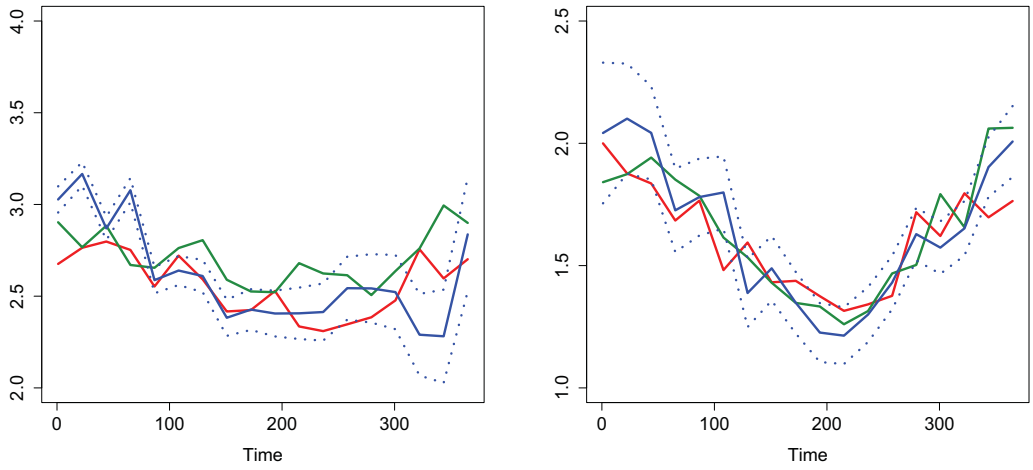

Figure 5.9: 0.01 - expectile curves for London (left) and Rome (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the latest 12 years expectile. expectiles.R

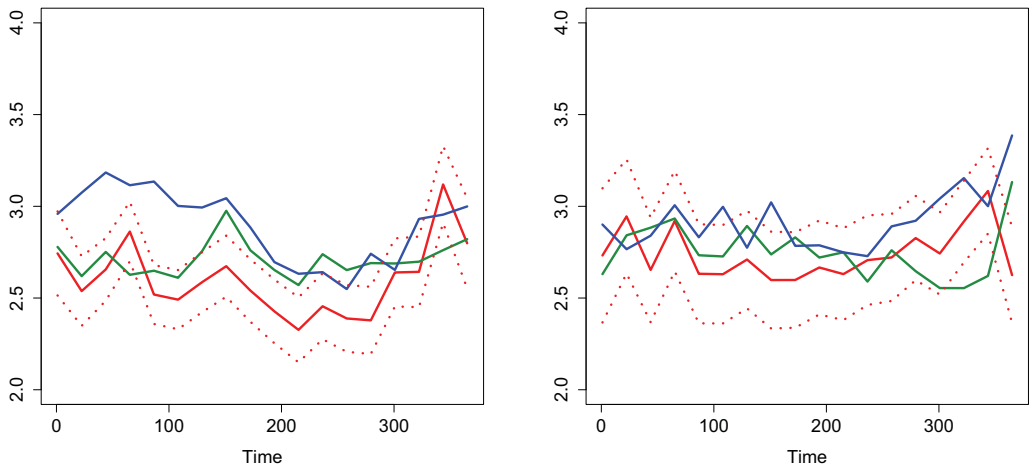

Figure 5.10: 0.01 - expectile curves for Paris (left) and Amsterdam (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the first 12 years expectile. expectiles.R

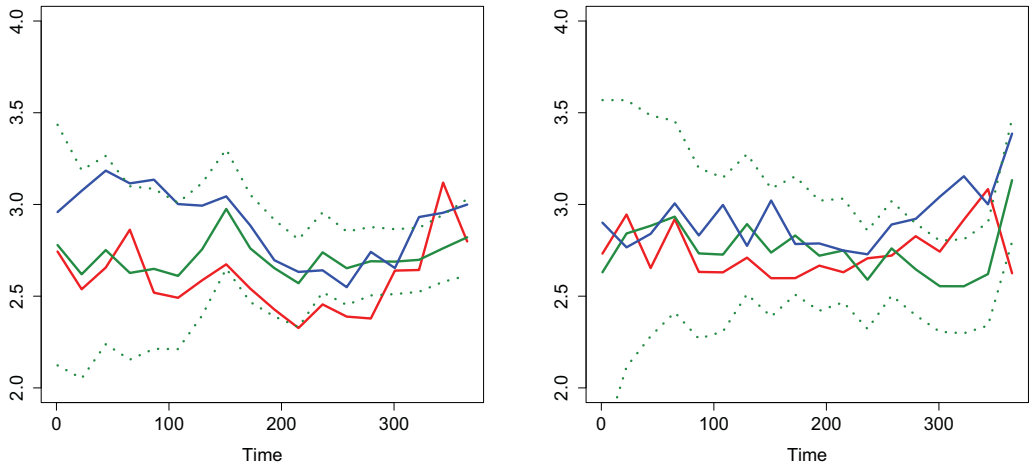

Figure 5.11: 0.01 - expectile curves for Paris (left) and Amsterdam (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the second 12 years expectile. expectiles.R

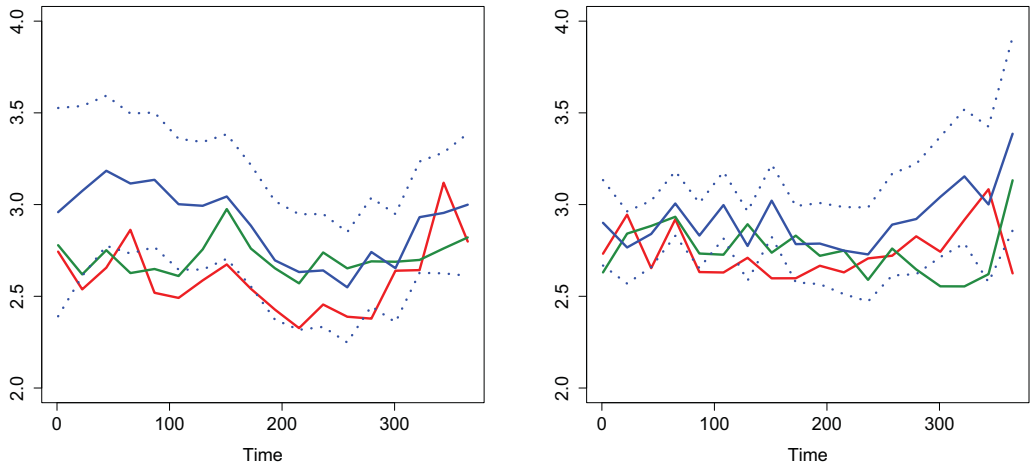

Figure 5.12: 0.01 - expectile curves for Paris (left) and Amsterdam (right) daily temperature variance from 1973 to 2008, red line for the first 12 years, green line for the second 12 years, blue line, for the latest 12 years, with the 5% - 95% confidence corridors for the latest 12 years expectile. expectiles.R

Figures 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11 and 5.12 depict the estimated expectile curves for London, Rome, Paris and Amsterdam. In each figure the red line corresponds to the expectile curve for the first pieriod $1973 - 1984$, the green line to the period $1985 - 1996$ and the blue line to the $1997 - 2008$ period. For the sake of brevity the fitted $5\% - 95\%$ confidence corridor is displayed only for one expectile curve. For 0.9 expectile, we attribute the figures 5.1, 5.2, 5.3, 5.4, 5.5 and 5.6 to extreme temperatures - squared model residuals observed within the sample. The 0.01 expectile shown in figures 5.7, 5.8, 5.9, 5.10, 5.11 and 5.12 correspond to the smallest squared residuals - the observations well explained by the model given in 2.7. It is worth to notice here, that as reported in Guo and Härdle (2012) expectiles are more robust approach for very high and very low τ, equation (5.1).

For each of the cities the expectile curves have similar spatial structure to Figure 3.11, the variance is significantly higher for the winter-fall period. The variance is at its heights in January, and the lowest variation is reported in July. The structure of fitted expectiles shows differences across the cities.

The most interesting findings coming from the fitted expectiles are the differences within each 12-year period. The extreme temperatures revealed by 0.9 expectiles differ significantly over each subsample. We report that the expectile lines fitted for different periods are not located within the $5\% - 95\%$ confidence corridor of the other curve. Moreover, except London, the values of the expectile curves grow with time, the red line corresponding to $1973 - 1985$ period lays below the green and blue curves; the blue line attributed to $1997 - 2008$ period is significantly higher than others. These results might be explained by the global warming effect, reported in seasonal function and more widely in literature. These findings hold across 3 out of 4 cities taken into consideration. The exception of London might be explained by the extensive human activities and industries localized in London area within $1973 - 1985$ and its influence

on the temperature similarly to findings reported in Guo and Härdle (2012).

The study of the low, 0.01 expectiles do not reveal significant differences within different periods. All of the fitted expectile curves differ significantly over each subsample. We report that the expectile lines fitted for different periods are not located within the $5\% - 95\%$ confidence corridors. Moreover there is no seasonal pattern, curves do not fluctuate much within a year. The only exception is Rome, what might be attributed to the higher temperatures reported, in comparison to Amsterdam, Paris and London.

6 Quantile Curves

Shifts of variance of daily average temperatures not only can be detected with the application of expectile functions but also with the help of local quantile regression. We follow Härdle et al. (2011b) proposal, of an adaptive local quantile regression algorithm. It was shown, that quantiles curves are good indicators for finding shifts in variance of local temperature residuals.

The τth quantile curve is given by the following formula:

$$Y_i = l(X_i) + \varepsilon_t \tag{6.1}$$

with $P(\varepsilon_t > 0) = \tau$ and $l(x)$, the conditional quantile function $F^{-1}_{Y|x}(\tau)$ which can be approximated by a polynomial. Y_i and X_i, with $i = 1, \ldots, n$, are independent random variables and $\tau \in (0,1)$.

The quantile regression, similarly to the expectiles, is applied to the daily squared residuals over each 12 year period after taking out seasonal and AR effects. X_i, with $i = 1, \ldots, n$ are the days of each year and Y_i are the daily squared residuals within each 12-year period.

Figures 6.1, 6.2, 6.3, 6.4 display the estimate quantile curves (red lines) and daily average squared residuals (blue dots) for each 12 year subsample as well as for all 36 years for the cities of London, Rome, Paris and Amsterdam. The lower red line corresponds to the 0.05 quantile curve, the middle one to the 0.5 quantile curve and the upper one to the 0.95 quantile curve. The 0.05 and 0.5 quantile curves are not varying significantly in comparison to 0.95 quantile curves for all cities. Therefore, we focus our analysis on understanding the 0.95 quantiles which correspond to the greatest values of daily average squared residuals.

For London, Figure 6.1 we observe the very interesting phenomenon that although for the first 24 years the variances in summer months are not volatile, for latest 12 years there is an upward tendency which reaches its highest peak at the end of august. For the second 12 years we observe increase of the variances for the winter months.

For Rome, Figure 6.2, we have the same yearly scheme of the quantile curves for each 12 year subperiod. The variance is higher in the beginning of the fall until the end of winter and lower from march until the end of summer. The same conclusion is remarked for the latest 12 year period of Paris, Figure 6.3, while for the previous years the variances are more volatile. For Amsterdam, Figure 6.4 depicts lower quantile levels during the spring and summer period and higher for fall and winter.

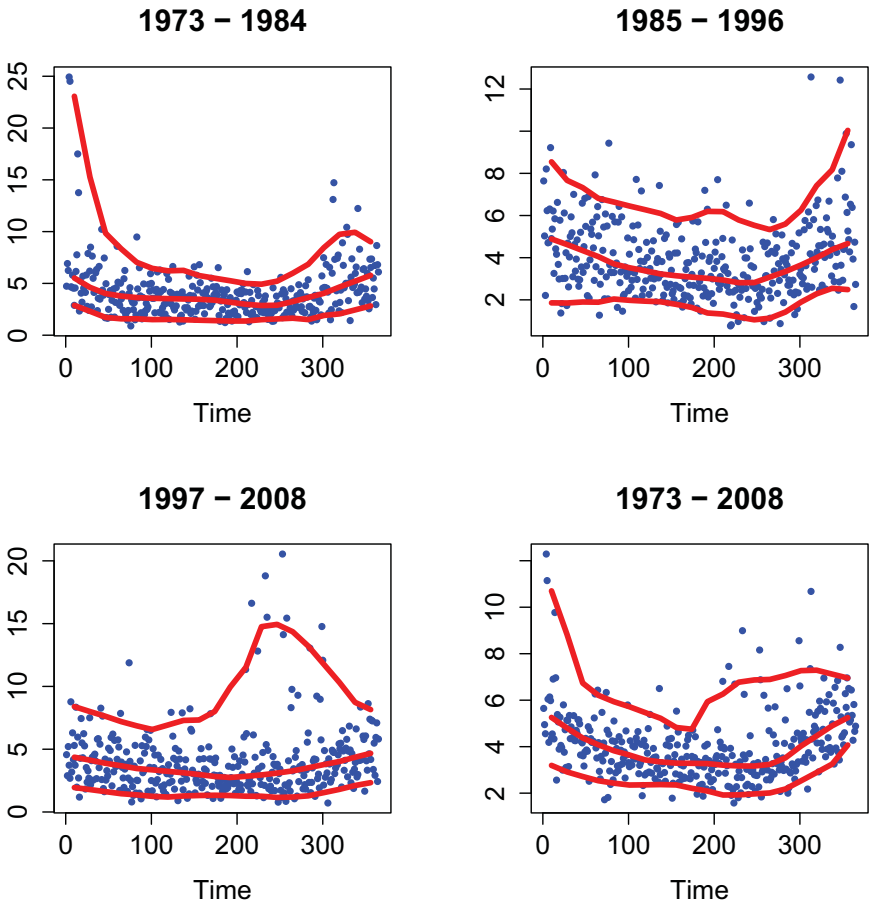

Figure 6.1: 0.05, 0.5 and 0.95 - quantile curves for London. quantiles.R

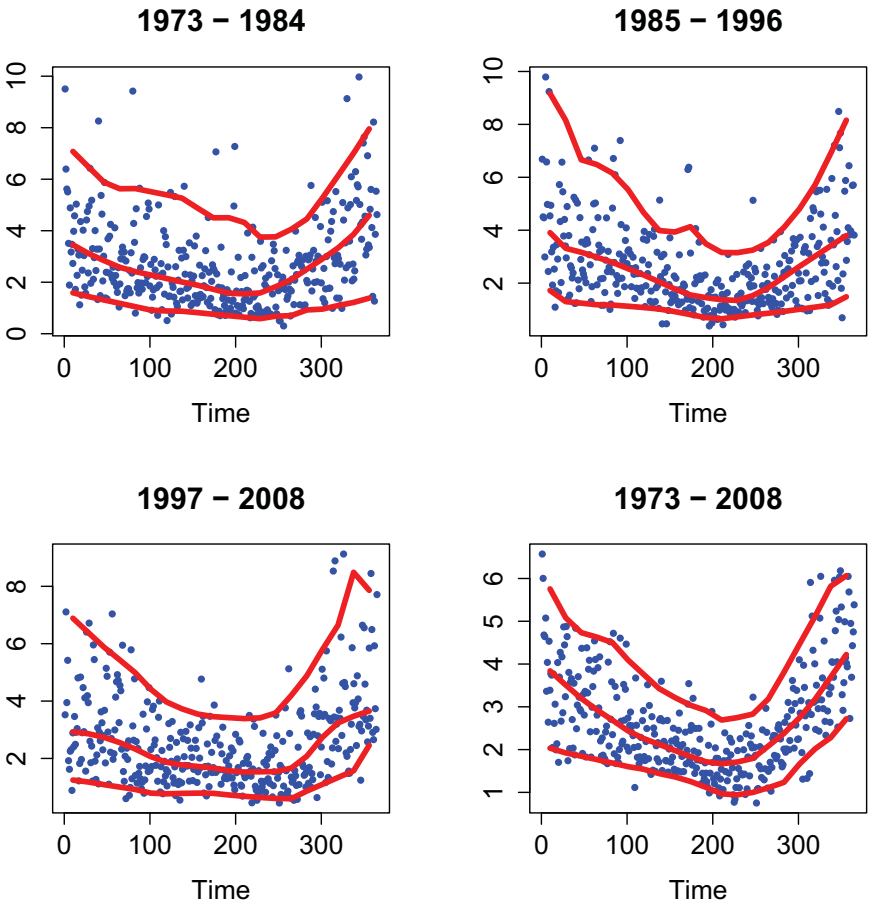

Figure 6.2: 0.05, 0.5 and 0.95 - quantile curves for Rome.Q quantiles.R

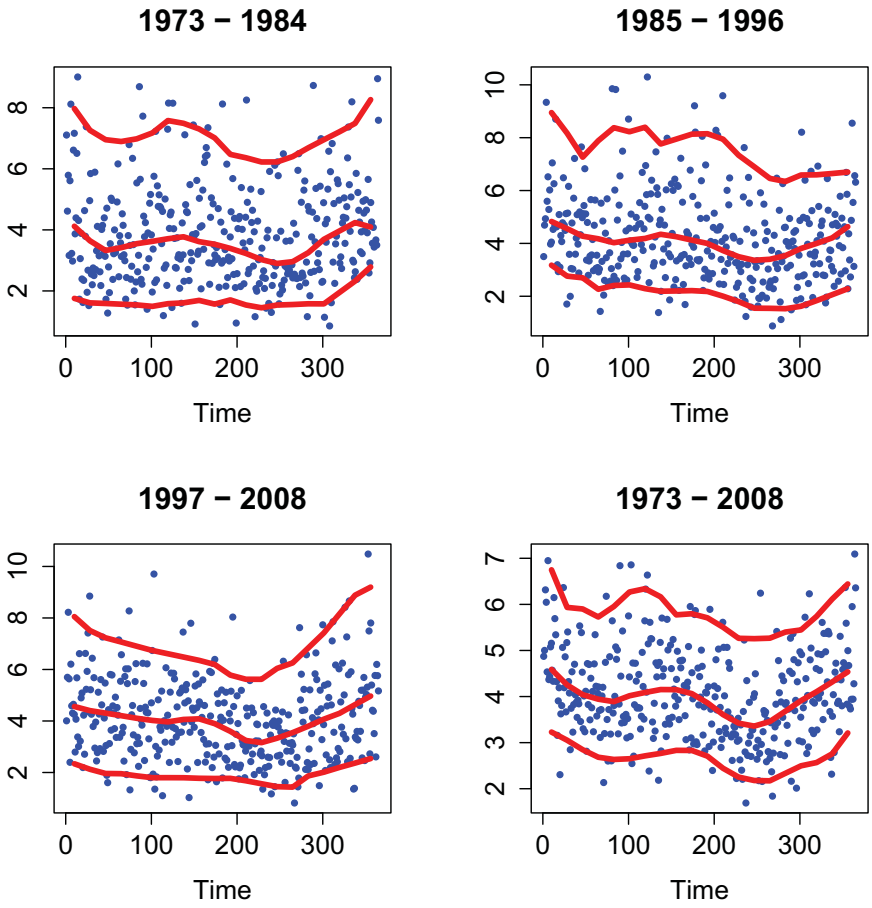

Figure 6.3: 0.05, 0.5 and 0.95 - quantile curves for Paris. quantiles.R

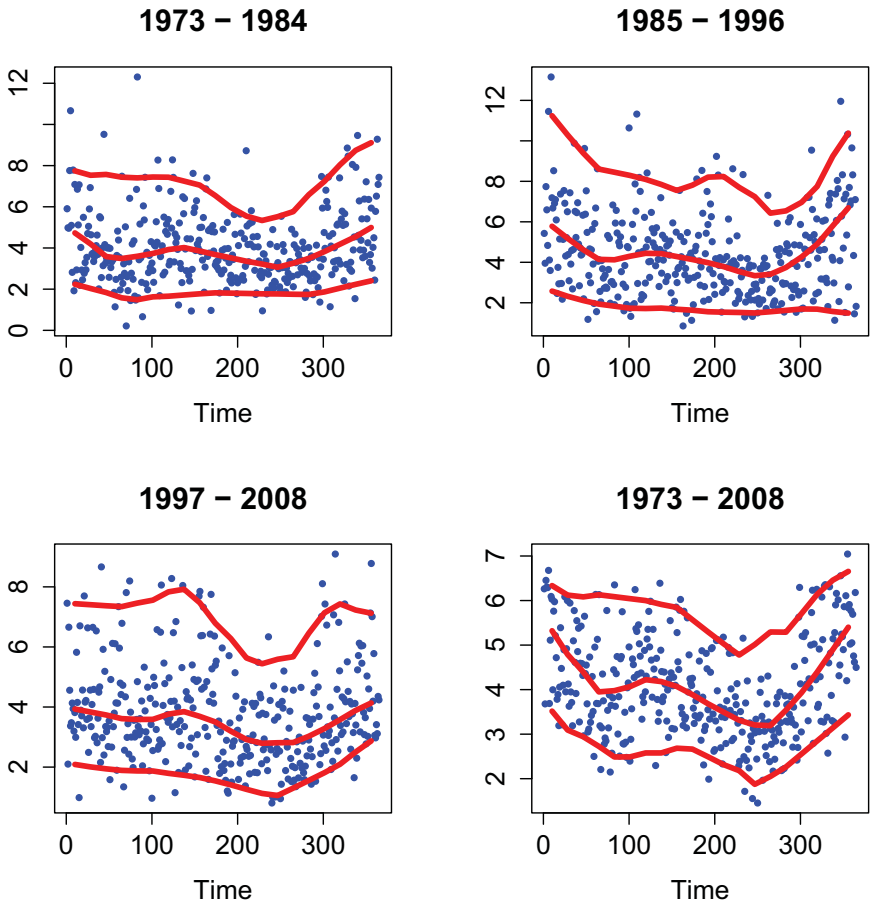

Figure 6.4: 0.05, 0.5 and 0.95 - quantile curves for Amsterdam. quantiles.R

7 Conclusion

This work analyses the daily average temperatures of the industrial Blue Banana European area on example of: London, Rome, Paris and Amsterdam. The sample covers the period from 1973 to 2008. At the beginning data were detrended and deseasonalised. Two methods were used in order to deseasonalize the data: the Local Linear Regression and the least squares method. For further analysis we applied the Lasso regression. This helped us to avoid overspecification and improve the selection of the seasonal component of the applied temperature model. Secondly, the higher order autoregressive process was applied to deseasonalized time series with seasonal volatility.

Four models proposed in literature were estimated: the Local Linear Regression (LLR) model, a model of Fourier series, an additive model of Fourier and GARCH terms and a multiplicative model of Fourier and GARCH terms. The seasonal variance of the residuals is calibrated with these four models. Next, we standardized the deseasonalized and detrended time series with the seasonal volatility obtained by the latter models The (standardized) residuals were investigated and compared. Summary statistics and normality tests were shown. We conclude that the model with a multiplicative structure of Fourier series and GARCH effect outperforms the other competitors in sample fit. The standardized residuals derived by the multiplicative model of Fourier and GARCH terms are close to the normal distribution (at significance level $\alpha = 0.05$, the normality hypothesis was not rejected). For that reason we performed extensive model validation of the latter one. Model validation follows the approach proposed by Benth and Saltyte Benth (2011). The out-of-sample analysis confirmed the goodness of fit. In order to evaluate our short term (one day) forecasting performance, the root mean squared prediction error and the mean absolute error were derived. The best fit among cities is achieved for Rome.

In the next step we applied expectile and quantile regression. The methods were applied to squared daily residuals derived from the deseasonalized and detrended daily temperetures over analyzed period. In order to study possible evolution of the variances of the temperature, the sample $1973-2008$ was divided into three subperiods $1973-1984$, $1985-1996$ and $1997-2008$. The 0.01 and 0.9 expectiles and 0.05, 0.5 (median) and 0.95 quantile curves were derived for all four cities under investigation. The significant shift in the extreme temperature observations, reflected by the 0.9 expectiles is found for three cities: Rome, Paris and Amsterdam. We observed, that the 0.9 expectile curve moves higher with time. The 0.01 expectile curve which is attributed to the observations well explained by the model do not show this phenomenon. The quantile regression confirms those findings. Additionally we reported increment in the variance between fall and winter, and relatively lower level behaviour of the temperature variances during the summer.

We hope that our extended analysis is helpful in understanding the temperature dynamics. We found some interesting phenomena which might be attributed to the global warming effect. For that reason it is worth to study the spatial structure of the temperature time evolution over different cities and locations. The applied models, especially the multiplicative model of Fourier

and GARCH terms are useful tools for further research.

Bibliography

Benth, F. ., Härdle, W. K., and López Cabrera, B. (2011). *Pricing Asian temperature risk in Statistical Tools for Finance and Insurance 2nd. edition (Cizek, Härdle and Weron, eds.)*. Springer Verlag Heidelberg.

Benth, F. E. and Saltyte Benth, J. (2011). A critical view on temperature modelling for application in weather derivatives markets. *Energy Economics*. Available on `http://dx.doi.org/10.1016/j.eneco.2011.09.012`. Version: October 8, 2011.

Benth, F. E., Saltyte Benth, J., and Koekebakker, S. (2007). Putting a price on temperature. *Scadinavian Journal of Statistics*, 34:746–767.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Clarendon Press, Oxford.

Campbell, S. D. and Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of American Statistical Association*, 100(469):6–16.

Guo, M. and Härdle, W. K. (2012). Simultaneous confidence bands for expectile functions. *Advances in Statistical Analysis*. Available on `http://dx.doi.org/10.1007/s10182-011-0182-1`. Version: December 3, 2011.

Härdle, W. K. and López Cabrera, B. (2011). The implied market price of weather risk. *Applied Mathematical Finance*. Available on `http://dx.doi.org/10.1080/1350486X.2011.591170`. Version: October 17, 2011.

Härdle, W. K., López Cabrera, B., Okhrin, O., and Wang, W. (2011a). Localising temperature risk. *Working paper, SFB 649 Humboldt Universität zu Berlin*.

Härdle, W. K., Spokoiny, V., and Wang, W. (2011b). Local quantile regression. *Working paper, SFB 649 Humboldt Universität zu Berlin*.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76:299–307.

Hyndman, R. J. and Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22.

Song, S., Härdle, W. K., and Ritov, Y. (2010). High dimensional nonstationary time series modelling with generalized dynamic semiparametric factor model. *Working paper, SFB 649 Humboldt Universität zu Berlin*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3):273–292.

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated sources. Where I have consulted the published work of others, in any form (e.g. ideas, equations, figures, text, tables), this is always explicitly attributed.

Berlin, 11 January 2012 Zografia Anastasiadou

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, 11 Januar 2012 Zografia Anastasiadou